

UCLA

UCLA Previously Published Works

Title

Developing an electronic health record measure of low-value esophagogastroduodenoscopy for GERD at a large academic health system.

Permalink

<https://escholarship.org/uc/item/9gg5t019>

Journal

BMJ Open Quality, 12(4)

Authors

Reynolds, Courtney

Nair, Vishnu

Villaflores, Chad

et al.

Publication Date

2023-12-22

DOI



10.1136/bmjopen-2023-002363

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at <https://creativecommons.org/licenses/by-nc/4.0/>

Peer reviewed

BMJ Open Quality **Developing an electronic health record measure of low-value esophagogastroduodenoscopy for GERD at a large academic health system**

Courtney A Reynolds ¹, Vishnu Nair,² Chad Villaflores,¹ Katherine Dominguez,¹ Julia Cave Arbanas,¹ Madeline Treasure,¹ Samuel Skootsky,³ Chi-Hong Tseng,¹ Catherine Sarkisian ^{3,4}, Arpan Patel,⁵ Kevin Ghassemi,⁵ A Mark Fendrick,⁶ Folasade P May,⁷ John N Mafi^{3,8}

To cite: Reynolds CA, Nair V, Villaflores C, *et al.* Developing an electronic health record measure of low-value esophagogastroduodenoscopy for GERD at a large academic health system. *BMJ Open Quality* 2023;**12**:e002363. doi:10.1136/bmjopen-2023-002363

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2023-002363>).

Received 31 March 2023
Accepted 3 December 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr John N Mafi;
jmafi@mednet.ucla.edu

ABSTRACT

Objectives Low-value esophagogastroduodenoscopies (EGDs) for uncomplicated gastro-oesophageal reflux disease (GERD) can harm patients and raise patient and payer costs. We developed an electronic health record (EHR) 'eMeasure' to detect low-value EGDs.

Design Retrospective cohort of 518 adult patients diagnosed with GERD who underwent initial EGD between 1 January 2019 and 31 December 2019.

Setting Outpatient primary care and gastroenterology clinics at a large, urban, academic health centre.

Participants Adult primary care patients at the University of California Los Angeles who underwent initial EGD for GERD in 2019.

Main outcome measures EGD appropriateness criteria were based on the American College of Gastroenterology 2012 guidelines. An initial EGD was considered low-value if it lacked a documented guideline-based indication, including alarm symptoms (eg, iron-deficiency anaemia); failure of an 8-week proton pump inhibitor trial or elevated Barrett's oesophagus risk. We performed manual chart review on a random sample of 204 patients as a gold standard of the eMeasure's validity. We estimated EGD costs using Medicare physician and facility fee rates.

Results Among 518 initial EGDs performed (mean age 53 years; 54% female), the eMeasure identified 81 (16%) as low-value. The eMeasure's sensitivity was 42% (95% CI 22 to 61) and specificity was 93% (95% CI 89 to 96). Stratifying across clinics, 62 (74.6%) low-value EGDs originated from 2 (12.5%) out of 16 clinics. Total cost for 81 low-value EGDs was approximately US\$75 573, including US\$14 985 in patients' out-of-pocket costs.

Conclusions We developed a highly specific eMeasure that showed that low-value EGDs occurred frequently in our healthcare system and were concentrated in a minority of clinics. These results can inform future QI efforts at our institution, such as best practice alerts for the ordering physician. Moreover, this open-source eMeasure has a much broader potential impact, as it can be integrated into any EHR and improve medical decision-making at the point of care.

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ Unnecessary, or low-value, EGDs can lead to direct patient harm, overburden the healthcare system with additional procedures and increase patient and payer costs.
- ⇒ Quality improvement efforts would benefit from the development of coding-based tools that can replace manual chart review.

WHAT THIS STUDY ADDS

- ⇒ We developed an electronic health record-based coding tool (eMeasure), which allows the prevalence of low-value EGD to be assessed without relying on chart review and can be used serially during quality improvement efforts.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ We plan to conduct a quality improvement project within our own healthcare system to reduce low-value EGDs.
- ⇒ Broadly, our eMeasure tool will be made available for widespread use, allowing other researchers and healthcare systems to conduct their own assessments and quality improvement efforts.

INTRODUCTION

Low-value medical care comprises testing, medication or procedures that offer no net benefit to patients in specific clinical scenarios, potentially causing patient harm and increased costs.¹⁻⁶ Given the potential for patient harm, ensuring the appropriate use of upper endoscopy is a high priority in the US healthcare system.^{7 8} Physicians perform an estimated 6.1 million esophagogastroduodenoscopies (EGDs) in the USA each year, with gastro-oesophageal reflux disease (GERD) as the most common indication.⁸ Adverse events occur in approximately 1 in 5000 procedures but can be serious, including bleeding, perforation, infection

and sedation-related complications.⁹ Despite these well-established risks, many EGDs lack an evidence-based indication. Global estimates of low-value EGD vary widely, but are as high as 36%–77%, suggesting an opportunity for improving the quality and value of care.^{10 11} A high volume of unnecessary EGDs may contribute to delays in scheduling of both clinically indicated EGDs and other procedures such as screening colonoscopy due to limited clinician time and procedure rooms. Low-value EGDs can overburden the health system and raise costs for patients and payers, in addition to their potential for direct patient harm.^{12 13}

Several quality improvement (QI) interventions have reduced low-value EGDs for dyspepsia, including workshops for physicians to review established practice guidelines and direct feedback to physicians on the appropriateness of their EGD referrals.¹³ However, no interventions to our knowledge have addressed low-value EGDs for GERD, despite GERD's status as the second most common gastrointestinal diagnosis in ambulatory settings, and the most common reason for EGD referral.^{8 13} The use of electronic health record (EHR) data to assess low-value EGDs has the potential to provide robust, reliable estimates that can determine the extent of the problem at individual sites and deliver serial clinician feedback during intervention efforts. While EHR data are often accessed by direct methods such as chart review, use of an electronic measure (eMeasure) can automate this process. An eMeasure is a standardised performance measure, which in this case tracks the quality of a particular healthcare service (ie, EGD), by extracting and analysing data from the EHR according to preprogrammed coding logic.

To the best of our knowledge, we developed and validated the first open-source eMeasure of low-value EGD in the initial management of GERD. The eMeasure builds on our experience developing a successful eMeasure of low-value colorectal cancer screening.¹⁴ Ultimately, this new eMeasure could be implemented by clinicians across the USA to identify the cost and prevalence of low-value EGD, track the effectiveness of future QI interventions, with the potential to improve the value of care for millions of Americans. Using the eMeasure, we also sought to characterise the prevalence of low-value care for initial EGD among adults with GERD at the University of California Los Angeles (UCLA) Health, a large academic health system across multiple clinical settings.

MATERIALS AND METHODS

Population

We identified the population of UCLA primary care patients aged ≥ 18 years who underwent an initial EGD between 1 January 2019 and 31 December 2019, with an International Classification of Diseases, 10th revision (ICD-10) diagnosis code of GERD or heartburn associated with their procedure or in the preceding 12 months. UCLA primary care patients are defined as having two

or more UCLA primary care physician (PCP) visits in the past 36 months, one or more UCLA preventive care PCP visits in the past 12 months or current membership in UCLA medical group's insurance plan. We excluded patients who had a previous EGD identified in our system within the past 36 months. We also excluded patients with known Barrett's oesophagus (BE), gastrointestinal malignancy, prior or planned bariatric surgery or chronic liver disease, as these conditions are important indications for both initial and serial EGDs distinct from indications for patients with GERD (a complete list of diagnostic codes has been submitted, along with our statistical coding language which is located in online supplemental files 1 and 2).

Development and validation of an electronic health record measure to assess EGD appropriateness

Building on our prior published methods,^{14 15} we developed an eMeasure to identify initial EGDs in the population described above and to characterise those EGDs as either high-value or low-value, based on ICD-10 coding data available in the EHR (Epic). We defined an EGD as 'initial' if it was the first EGD to occur within a 3-year period. This timeline is consistent with prior studies of repeat EGD, which typically suggest that after a 3-year period, a repeat EGD may be reasonable, as the patient's clinical situation may have changed enough to warrant repeat EGD regardless of the initial EGD's findings (or lack thereof).^{7 11} Two board-certified gastroenterologists and two board-certified general internists defined multidisciplinary EGD appropriateness criteria using the evidence-based 2012 American College of Gastroenterology guidelines for GERD management.¹⁶ Manual chart review was then conducted for a random sample of 204 patients to validate these criteria in identifying low-value EGD. To address the problem of incomplete and non-specific diagnosis coding in routinely collected EHR data, we explicitly designed the eMeasure to use the broadest possible interpretation of codes to capture indications (eg, anaemia was treated as iron deficiency anaemia, and unspecified weight loss was treated as unintentional, clinically significant weight loss) to achieve our prespecified goal of $\geq 90\%$ specificity. We chose to maximise specificity over sensitivity to minimise false positives that would mislabel appropriate EGDs as low-value, and potentially prevent appropriate care should the eMeasure be used to inform a best practice alert in the future. We chose the value of 90% specificity in particular based on decision analysis literature on reliability and validity, which cite this level of specificity as a marker of good test performance.^{17 18} Our high-specificity approach also ensured an eMeasure that would be as generous as possible to clinicians, as we are aware of the limited ability of routine billing codes to convey all the clinicians' concerns and reasoning about a given patient. In doing so, we believed our measure would have the greatest credibility among clinicians during future QI efforts.

Table 1 Three different definitions of elevated risk of BE¹⁹⁻²¹

Strictest criteria ²¹	Intermediate criteria ^{16 20}	Least strict criteria ^{16 20}
Male gender plus GERD symptoms >5 years	Any four risk factors listed under the strictest criteria	Any three risk factors listed under the strictest criteria
Plus two or more of the following:		
Age >50 years		
White		
BMI >25 kg/m ²		
Current or past smoking		
Family history of BE or EAC		

BE, Barrett's oesophagus; BMI, body mass index; EAC, oesophageal adenocarcinoma; GERD, gastro-oesophageal reflux disease.

An EGD was considered low-value if none of the following criteria were met: (1) alarm symptom present (eg, diagnostic codes of gastrointestinal bleeding, iron deficiency anaemia, weight loss, etc); (2) completion of 8 weeks of proton pump inhibitor (PPI) therapy or (3) elevated risk of Barrett's oesophagus (BE). Since no single consensus definition for elevated BE risk exists, we performed sensitivity analyses by testing three different definitions that were based on available guidelines, in order to determine if use of different guidelines affected our results.^{16 19-21} Our prespecified goal was to maximise eMeasure specificity; therefore we used the most inclusive definition of elevated risk, for which patients need to have three risk factors, such as smoking, family history of BE or obesity (table 1). PPI use was obtained from prescription data and did not include over-the-counter use.

We performed manual chart review to assess the validity of our eMeasure. After establishing high inter-rater reliability on a testing set (see 'Results' section for details), we then reviewed 204 randomly selected charts and calculated the sensitivity and specificity of the measure. Results from manual chart review were treated as the gold standard results, and those from the eMeasure were the test results. For example, a true positive (low-value EGD) was identified as such by both the eMeasure and by chart review, while a false positive was identified as low-value by the eMeasure but not by chart review. Sensitivity is the proportion of people with low-value EGD who are correctly identified by the eMeasure, or the number of true positives divided by the sum of true positives and false negatives. Similarly, specificity is the proportion of people without low-value EGD who are correctly identified by the eMeasure, or the number of true negatives divided by the sum of true negatives and false positives. Since sensitivity and specificity are proportions, their 95% CIs can be calculated using a binomial distribution. We collected patient demographics such as age, sex, race/ethnicity and estimated income using 5-digit zip code and comorbidities (table 2). We calculated low-value

EGD costs using 2022 Medicare physician and facility fee rates (US\$933: CPT code 43235), which also provides estimates for patient out-of-pocket costs.²² We stratified the frequency of EGDs by clinic site, based on where the originating EGD referral order took place, to better inform future QI initiatives.

Statistical analysis

Our statistical power calculations revealed that 204 medical charts would have approximately 80% power to demonstrate our prespecified goal of an eMeasure specificity of 90% or higher. For this power calculation, we assumed a low-value EGD prevalence of 20% and specificity of 95% based on preliminary estimates from our initial exploratory chart review of 20 random cases. We reviewed an additional 15 random cases to test our inter-rater reliability prior to validating the eMeasure. These 15 cases were not included in the 204 charts randomly selected for review for eMeasure validation. We performed two-tailed t-tests and χ^2 tests to assess the relationship between various demographic and socioeconomic characteristics and EGD overuse, with a p value <0.05 as statistically significant. SAS V.9.4 was used for these comparison tests; all other analyses were conducted using SQL Server Management Studio V.14.0.17289.0.

Patient and public involvement

The public and patients were not involved in the design, conduct, reporting or dissemination plans of our research.

RESULTS

Among 21 437 adults with GERD or heartburn, 518 (2.4%) underwent an initial EGD during the study period (figure 1). Average age was 55 years (SD=14), and 54% of the population was female. For race and ethnicity, 61% of patients identified as white, 8% Asian, 3% black, 14% other and 14% had missing race/ethnicity information (table 2).

Measure performance and validity

After reviewing the guideline-based criteria, the blinded physician reviewers independently agreed >90% of the time whether the EGD was appropriate in a training set (n=15), demonstrating excellent inter-reviewer reliability. Each of the three reviewers then independently reviewed 68 randomly selected EGD cases for a total of 204 unique cases. This manual chart review of 204 cases represented the gold standard evaluation of low-value EGD to which the eMeasure was compared. When compared with physician chart review, the eMeasure had an overall specificity of 93% (95% CI 89 to 96) and sensitivity of 42% (95% CI 22 to 61) for identifying low-value EGD.

Low-value EGD referrals for GERD in a large academic health system

Our eMeasure identified 518 EGDs among patients with GERD who did not have a documented prior EGD. Of these, 81 (16%) EGDs met our criteria for low-value care.

Table 2 Characteristics of patients referred for EGD, stratified by appropriateness (n=518)²⁸

	Potentially high-value (n=437) n (%)	Potentially low-value (n=81) n (%)	Total (n=518) n (%)	P value
Age (mean±SD)	56.4±14.3	49.9±15.7	55.4±14.7	<0.001
Gender				<0.001
Male	219 (50.1)	20 (24.7)	239 (46.1)	
Female	218 (49.9)	61 (75.3)	279 (53.9)	
Race				<0.001
White	280 (64.1)	34 (42.0)	314 (60.6)	
Asian	31 (7.1)	12 (14.8)	43 (8.3)	
Black	14 (3.2)	0 (0.0)	14 (2.7)	
AI/AN	0 (0.0)	1 (1.2)	1 (0.2)	
NH/PI	0 (0.0)	0 (0.0)	0 (0.0)	
Other	59 (13.5)	16 (19.8)	75 (14.5)	
Unknown	53 (12.1)	18 (22.2)	71 (13.7)	
Ethnicity				0.008
Not Hispanic	322 (73.7)	56 (69.1)	378 (73.0)	
Hispanic	58 (13.3)	5 (6.2)	63 (12.2)	
Unknown	57 (13.0)	20 (24.7)	77 (14.9)	
Annual household income (US\$) ²⁸				0.05
<50 000	12 (2.7)	4 (4.9)	16 (3.1)	
50 000–75 000	50 (11.4)	17 (21.0)	67 (12.9)	
75 000–100 000	95 (21.7)	16 (19.8)	111 (21.4)	
≥100 000	189 (43.2)	28 (34.6)	217 (41.9)	
Unknown	91 (20.8)	16 (19.8)	107 (20.7)	

P values reflect comparison of high-value versus low-value groups.

AI, American Indian; AN, Alaska Native; NH, Native Hawaiian; PI, Pacific Islander.

Low-value EGDs were associated with younger age (50 vs 56 years, $p<0.001$) and with female gender (75% vs 50% female, $p<0.001$) (see [table 2](#) for details). We also assessed the frequency of low-value EGD by clinic and found that approximately 61 of the 81 low-value EGDs (75%) originated from 2 clinics of 16 clinics total ([figure 2](#)). The clinics represented a mix of primary care and gastroenterology clinics; the two clinics with the most low-value EGDs were gastroenterology clinics. Combined, the 81 low-value EGDs had an estimated total cost of US\$75 573, including US\$14 985 in patient out-of-pocket costs (costs of US\$933 total per patient and US\$185 per patient out-of-pocket, respectively).

Sensitivity analysis of applying various definitions of high risk for BE

In total, there were 460 out of 518 EGDs performed on patients at elevated risk of BE, using the most inclusive definition of elevated risk. We found that using any of our three guideline-based definitions of elevated risk for BE did not substantially impact our results. For instance, the specificity associated with the three models was as follows, in order from most restrictive to most inclusive model: model 1 87% (95% CI 82% to 93%), model 2 87% (95% CI 82% to 92%) and model 3 92% (95% CI 88% to 96%).

The sensitivity for each model was also highly similar: model 1 38% (95% CI 25% to 50%), model 2 37% (95% CI 22% to 50%) and model 3 39% (95% CI 19% to 50%). In accordance with our prespecified plan to maximise eMeasure specificity, our final analysis used the most inclusive and broad definition of BE risk, requiring only three known risk factors ([table 1](#)).

CONCLUSIONS

We developed a highly specific eMeasure to identify low-value EGD among patients with GERD and heartburn. The use of our EGD metric is a novel progression from prior studies that have drawn attention to the issue of low-value care in endoscopy. Most studies have focused on colonoscopy, primarily for screening.^{6 7 15} A notable, large-scale study of low-value EGD was completed in the Veteran's Affairs hospital system but was restricted to repeat EGD only.¹¹ The majority of studies, whether on colonoscopy or EGDs, also rely heavily on chart review. To our knowledge, this is the first study to introduce a validated, replicable, automated EHR-based approach to identifying and tracking low-value EGDs in the initial management of GERD.

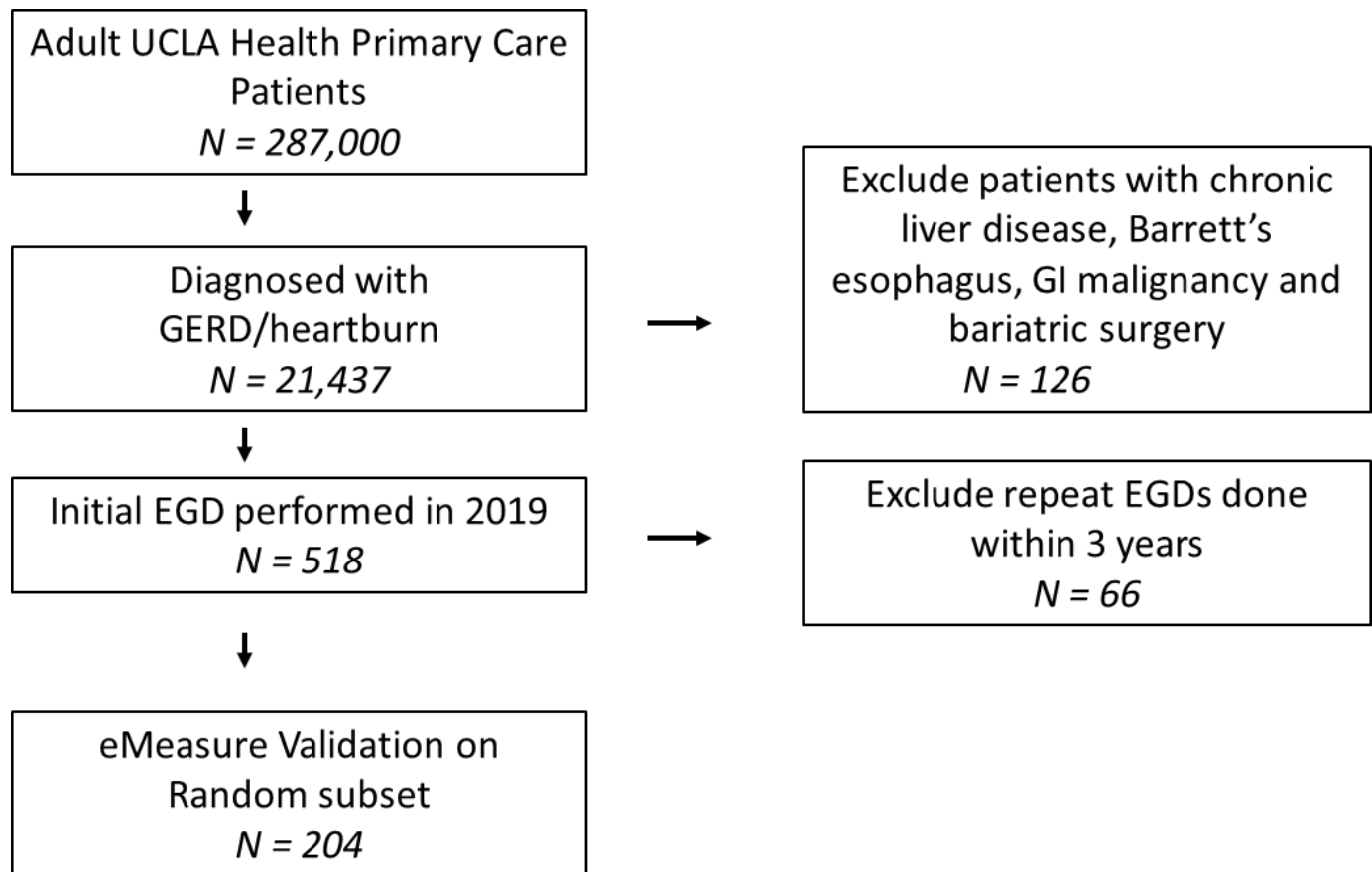


Figure 1 Patient population. EGD, esophagogastroduodenoscopy; GERD, gastro-oesophageal reflux disease; GI, gastrointestinal; UCLA, University of California Los Angeles.

Our eMeasure results suggest that up to 4 in every 25 EGDs are low-value, exposing patients to unnecessary procedure-related risks and avoidable out-of-pocket costs. As we designed our measure with high specificity, the actual incidence of low-value EGDs is likely to be even higher than our estimates, highlighting the importance of developing QI efforts to reduce low-value EGDs to lower spending while improving the quality of care.

Our results further suggest that referral patterns in a small number of clinics may provide a focus for future targeted intervention at our institution. Since low-value EGDs were especially high at two referral clinics, an initial, data-driven and non-judgemental gastroenterologist-led intervention could be focused on discussions with gastroenterologists at those two sites in the future.²³ To increase physician engagement, there would be open discussion of how to incorporate tactics such as allowing the clinics themselves to autonomously develop standardised local practice guidelines and implementing an EHR-based best practice alert influencing physician ordering patterns at the point of care. Our measure could then be used prospectively to serially assess low-value referrals and detect responses to the intervention compared with a control group.

In addition to use within our health system, our measure has the potential for broader impact. Our eMeasure is based solely on demographic data and

ICD-10 diagnosis codes that are present in any EHR, and thus could theoretically be adapted to function in any EHR. Use of the eMeasure can facilitate real-time, automated monitoring of low-value care at health systems across the USA, and therefore identify the unique drivers of low-value care at various institutions. The eMeasure could also be expanded to include a clinical decision tool for providers, such as a best practice alert on indications for EGD. In addition, our approach to developing a highly specific, EHR-based measure can be applied to topics beyond EGD and fields beyond gastroenterology. Low-value medical care can be successfully reduced using a respectful, data-driven approach that leverages non-judgmental communication.²³ Our measure was designed to maximise specificity, to give clinicians the ‘benefit of the doubt’ and thus could build greater trust in our results among referring physicians. An eMeasure that falsely classified an EGD as low-value when it was appropriate would rapidly lose credibility among frontline clinicians and might lead to unintended patient harm by disrupting medically necessary EGDs. The combination of a highly specific, credible eMeasure tool for tracking and clinician-centred interventions is likely to yield the optimum result in reducing low-value care in any field.

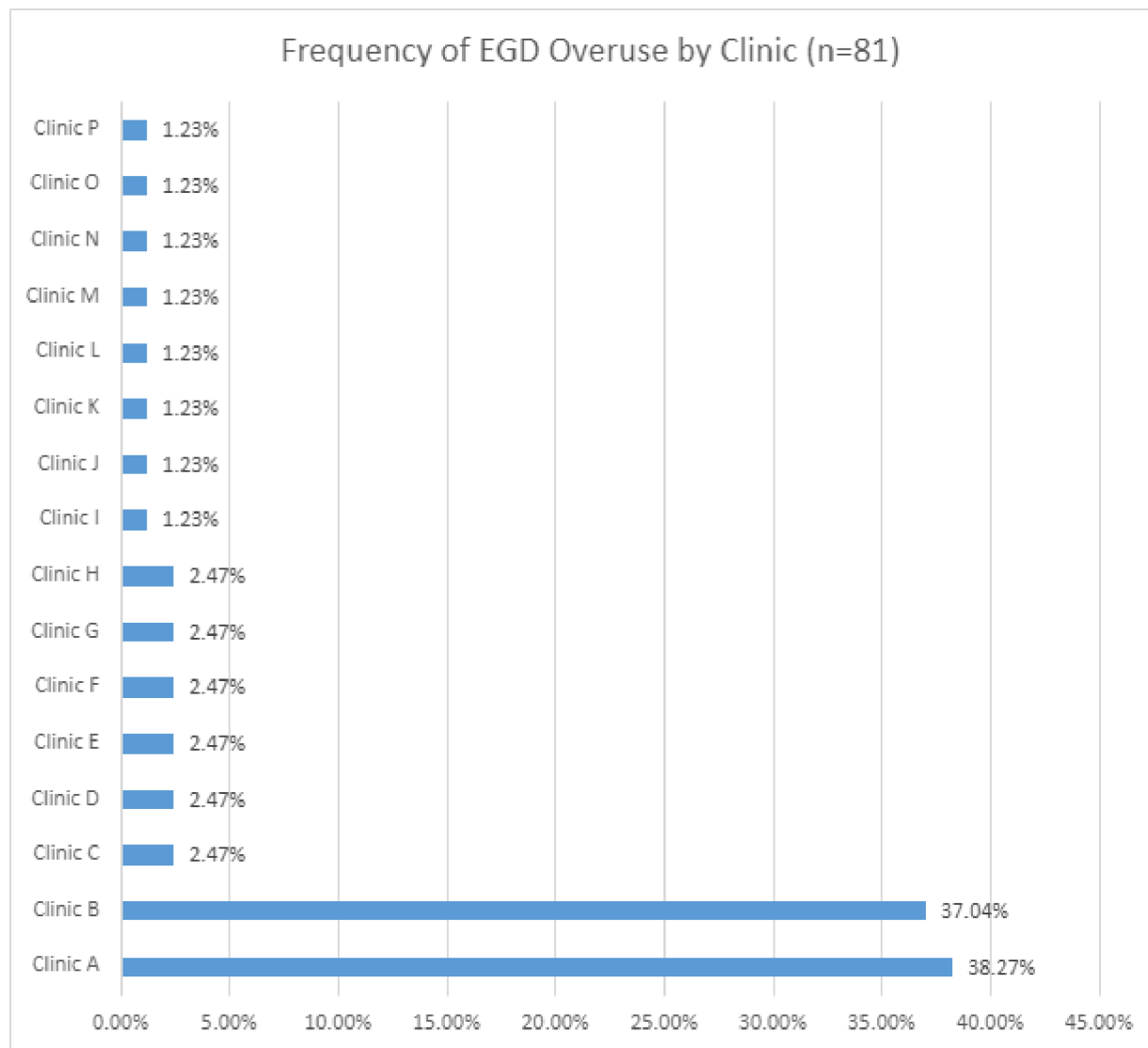


Figure 2 Esophagogastroduodenoscopies (EGDs) stratified by primary care and gastroenterology clinics.

Limitations

While this is the first study to our knowledge to use an eMeasure to identify low-value EGD, this was a single-centre retrospective analysis. Thus, our results may not necessarily generalise elsewhere, particularly for settings other than large, urban academic medical centres. Another important limitation is our reliance on routine coding for many aspects of the measure, including GERD diagnosis, PPI use and risk factors such as anaemia or smoking history. To compensate for inaccurate or non-specific coding, we used the most liberal definitions for these codes. For example, documentation of anaemia was deemed to be equivalent to iron deficiency anaemia, which would be a justification for EGD referral. In other words, to maximise specificity we favoured errors of commission over errors of omission when identifying indications for EGD.

As mentioned above, we intentionally designed our eMeasure to have high specificity, at the natural expense of sensitivity. In doing so, we reduce the chance that

the eMeasure—if used to drive a best practice alert or other intervention—would falsely label an appropriate EGD as low-value and thus present a barrier to patient care. In exchange, we allowed more potentially inappropriate EGDs. The cost associated with this approach is the facility and direct patient costs associated with any EGD (described in ‘Results’ section), as well as the cost of adverse events from the procedure itself. However, there would be potentially greater associated costs and harms if we were to take the opposite approach and design an eMeasure with high sensitivity and low specificity. If, for example, our sensitivity was 93% and specificity was 42% (reversing our actual numbers), the false positive rate would rise from 6 to as many as 49 cases in 100. Assuming that some of these cases were unable to undergo EGD in a timely manner as a result, there would be a delay in diagnosing serious conditions such as oesophageal cancer. For instance, the average cost associated with treatment of stage 1 oesophageal cancer is US\$73 595 vs US\$144 019 for stage 4.^{24 25} A delay in diagnosis could then result in

both patient harm and a potential doubling in costs for subsequent treatment.

To track PPI use, we were restricted to prescriptions and not able to capture over-the-counter PPI use. In addition, prescriptions for PPI may not have been filled, and adherence to filled prescriptions was not assessed. In the future, use of natural language processing may be able to address some of these challenges, specifically by improving eMeasure sensitivity without sacrificing specificity.²⁶

EGD appropriate use guidelines have differences, particularly on how to define elevated risk of BE.^{19–21} We conducted a sensitivity analysis on the impact of different guidelines for BE screening on EGD appropriateness but found that the impact on our measure's performance was small. This may reflect the fact that indications for BE screening are difficult to glean from coding, for example, family history is often included in notes rather than coded. Thus, the use of one guideline versus another did not alter our estimates substantially. Failure to capture patients at high risk for BE could result in an overestimation of inappropriate EGD referrals. To address this concern, we used the broadest and most inclusive definition of high BE risk possible, combined with rigorous chart review. Future work could examine whether incorporating differing weights of the eMeasure components such as BE risk can improve eMeasure performance.

Finally, costs were estimated using published Medicare physician fee schedule reimbursement rates and may differ for the patients in our study who are insured by Medicaid or commercial health plans. Nevertheless, US Medicare reimbursement rates are publicly available and easily reproducible by other researchers, and they serve as a standard benchmark for reimbursement rates set by all other payers.²⁷

In summary, we created a highly specific eMeasure for low-value EGD, at the intentional expense of measure sensitivity. Low-value EGDs occurred frequently and raised costs for patients and payers. Most low-value EGDs occurred in a small minority of clinics, which can inform future QI interventions at our institution. Furthermore, as our open-source eMeasure can function in any EHR, its use can identify targets for intervention at any institution and improve the value of physician decision-making at the point of care.

Author affiliations

¹Department of Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA

²Department of Medicine, Stanford University, Stanford, California, USA

³Division of General Internal Medicine and Health Services Research, David Geffen School of Medicine at UCLA, Los Angeles, California, USA

⁴Veterans' Administration Greater Los Angeles Healthcare System, Geriatric Research Education & Clinical Center (GRECC), Birmingham, Alabama, USA

⁵The Vatche and Tamar Manoukian Division of Digestive Diseases, Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California, USA

⁶Department of Medicine, University of Michigan Medical School, Ann Arbor, MI, USA

⁷UCLA Kaiser Permanente Center for Health Equity, Jonsson Comprehensive Cancer Center, Los Angeles, Calif, USA

⁸RAND Health, RAND Corporation, Santa Monica, California, USA

Acknowledgements We are thankful to Eric Esrailian, chair of the Vatche and Tamar Manoukian Division of Digestive Diseases, for his continued support and guidance throughout this project.

Contributors All authors met the uniform criteria for authorship. All authors have seen and reviewed the manuscript and take responsibility for its final content. JNM, guarantor.

Funding This work was supported by a NIH/NIA Beeson Emerging Leaders in Aging Research Career Development Award (grant K76AG064392-01A1) for JM.

Competing interests Dr Mafi reported grants from the National Institute on Aging (NIA) during the conduct of the study, as well as grants from Arnold Ventures and the Commonwealth Fund. Dr. Mafi previously received nonfinancial support from Milliman MedInsight and has provided unpaid consulting to Milliman MedInsight and AHRQ. Ms. Arbanas reported grants from the NIA during the conduct of the study. Dr Sarkisian reported grants from the National Institutes of Health (NIH) during the conduct of the study. No other disclosures were reported.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The UCLA Institutional Review Board (IRB) approved this retrospective cohort study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Data are available on reasonable request. Our source code, used to run the eMeasure, will be available 'open source' following publication.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Courtney A Reynolds <http://orcid.org/0000-0002-0298-1039>

Catherine Sarkisian <http://orcid.org/0000-0003-2768-2323>

REFERENCES

- Mafi JN, Russell K, Bortz BA, *et al*. High-volume health services contribute the most to unnecessary health spending. *Health Aff (Millwood)* 2017;36:1701–4.
- Morgan DJ, Dhruva SS, Wright SM, *et al*. Update on medical Overuse: a systematic review. *JAMA Intern Med* 2016;176:1687–92.
- Chan KS, Chang E, Nassery N, *et al*. The state of Overuse measurement: a critical review. *Med Care Res Rev* 2013;70:473–96.
- Newton EH, Zazzera EA, Van Moorsel G, *et al*. Undermeasuring Overuse—an examination of national clinical performance measures. *JAMA Intern Med* 2015;175:1709–11.
- Mafi JN, Parchman M. Low-value care: an intractable global problem with no quick fix. *BMJ Qual Saf* 2018;27:333–6.
- O'Sullivan JW, Albasri A, Nicholson BD, *et al*. Overtesting and Undertesting in primary care: a systematic review and meta-analysis. *BMJ Open* 2018;8:e018557.
- Leal C, Almeida N, Silva M, *et al*. Appropriateness of endoscopic procedures: A prospective, multicenter study. *GE Port J Gastroenterol* 2022;29:5–12.
- Peery AF, Crockett SD, Murphy CC, *et al*. Burden and cost of gastrointestinal, liver, and Pancreatic diseases in the United States: update 2018. *Gastroenterology* 2019;156:254–72.
- Wolfsen HC, Hemminger LL, Achem SR, *et al*. Complications of Endoscopy of the upper gastrointestinal tract: a single-center experience. *Mayo Clin Proc* 2004;79:1264–7.



- 10 Saini SD, Powell AA, Dominitz JA, *et al.* Developing and testing an electronic measure of screening colonoscopy Overuse in a large integrated Healthcare system. *J Gen Intern Med* 2016;31 Suppl 1(Suppl 1):53–60.
- 11 Rubenstein JH, Pohl H, Adams MA, *et al.* Overuse of repeat upper Endoscopy in the veterans health administration: A retrospective analysis. *Am J Gastroenterol* 2017;112:1678–85.
- 12 El-Serag HB, Sweet S, Winchester CC, *et al.* Update on the epidemiology of Gastro-Oesophageal reflux disease: a systematic review. *Gut* 2014;63:871–80.
- 13 de Jong JJ, Lantinga MA, Drenth JP. Prevention of Overuse: a view on upper gastrointestinal Endoscopy. *World J Gastroenterol* 2019;25:178–89.
- 14 Soroudi C, Mafi J, Myint A, *et al.* Leveraging electronic health records to measure low-value screening colonoscopy. *Am J Med* 2022;135:715–20.
- 15 Mafi JN, May FP, Kahn KL, *et al.* Low-value proton pump inhibitor prescriptions among older adults at a large academic health system. *J Am Geriatr Soc* 2019;67:2600–4.
- 16 Katz PO, Gerson LB, Vela MF. Guidelines for the diagnosis and management of gastroesophageal reflux disease. *Am J Gastroenterol* 2013;108:308–28.
- 17 Parikh R, Mathai A, Parikh S, *et al.* Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol* 2008;56:45–50.
- 18 Power M, Fell G, Wright M. Principles for high-quality, high-value testing. *Evid Based Med* 2013;18:5–10.
- 19 Gyawali CP, Fass R. Management of gastroesophageal reflux disease. *Gastroenterology* 2018;154:302–18.
- 20 Spechler SJ, Sharma P, Souza RF, *et al.* American Gastroenterological Association medical position statement on the management of Barrett's esophagus. *Gastroenterology* 2011;140:1084–91.
- 21 Shaheen NJ, Falk GW, Iyer PG, *et al.* ACG clinical guideline: diagnosis and management of Barrett's esophagus. *Am J Gastroenterol* 2016;111:30–50;
- 22 Medicare physician fee schedule look up tool. centers for Medicare and Medicaid services (CMS). Available: <https://www.cms.gov/apps/physician-fee-schedule/search/search-criteria.aspx> [Accessed 27 Sep 2021].
- 23 Beckman H, Mafi J, Bortz B. A 10-step program to successfully reduce low-value care. *Am J Manag Care* 2021;27:e208–13.
- 24 Reddy SR, Broder MS, Chang E, *et al.* Cost of cancer management by stage at diagnosis among Medicare beneficiaries. *Current Medical Research and Opinion* 2022;38:1285–94.
- 25 Kakushadze Z, Raghubanshi R, Yu W. Estimating cost savings from early cancer diagnosis. *Data* 2017;2:30.
- 26 McVay TR, Garrett G C, Celena B P, *et al.* Natural language processing accurately identifies Dysphagia indications for Esophagogastroduodenoscopy procedures in a large US integrated Healthcare system: implications for classifying Overuse and quality measurement. *AMIA Jt Summits Transl Sci Proc* 2019;2019:665–71.
- 27 Mafi JN, Godoy-Travieso P, Wei E, *et al.* Evaluation of an intervention to reduce low-value preoperative care for patients undergoing cataract surgery at a safety-Net health system. *JAMA Intern Med* 2019;179:648.
- 28 Kind AJH, Buckingham WR. Making neighborhood disadvantage Metrics accessible: the neighborhood Atlas. *N Engl J Med* 2018;378:2456–8.